# Post-radiochemotherapy PET radiomics in head and neck cancer - The influence of radiomics implementation on the reproducibility of local control tumor models

Bogowicz, Marta ; Leijenaar, Ralph T H ; Tanadini-Lang, Stephanie ; Riesterer, Oliver ; Pruschy, Martin ; Studer, Gabriela ; Unkelbach, Jan ; Guckenberger, Matthias ; Konukoglu, Ender ; Lambin, Philippe

Abstract: Purpose This study investigated an association of post-radiochemotherapy (RCT) PET radiomics with local tumor control in head and neck squamous cell carcinoma (HNSCC) and evaluated the models against two radiomics software implementations. Materials and methods 649 features, available in two radiomics implementations and based on the same definitions, were extracted from HNSCC primary tumor region in 18F-FDG PET scans 3 months post definitive RCT (training cohort n = 128, validation cohort n = 50) and compared using the intraclass correlation coefficient (ICC). Local recurrence models were trained, separately for both implementations, using principal component analysis (PCA) and the least absolute shrinkage and selection operator. The reproducibility of the concordance indexes (CI) in univariable Cox regression for features preselected in PCA and the final multivariable models was investigated using respective features from the other implementation. Results Only 80 PET radiomic features yielded ICC $> 0.8$ in the comparison between the implementations. The change of implementation caused high variability of CI in the univariable analysis. However, both final multivariable models performed equally well in the training and validation cohorts (CI $> 0.7$) independent of radiomics implementation. Conclusion The two post-RCT PET radiomic models, based on two different software implementations, were prognostic for local tumor control in HNSCC. However, 88% of the features was not reproducible between the implementations.

DOI: https://doi.org/10.1016/j.radonc.2017.10.023

Post-radiochemotherapy PET radiomics in head and neck cancer - The influence of radiomics implementation on the reproducibility of local control tumor models. Radiotherapy and Oncology, 125(3):385-391. DOI: https://doi.org/10.1016/j.radonc.2017.10.023

# *Post-radiochemotherapy PET radiomics in head and neck cancer - the influence of radiomics implementation on the reproducibility of local control tumor models*

[1]Marta Bogowicz, [2]Ralph TH Leijenaar, [1]Stephanie Tanadini-Lang, [1]Oliver Riesterer, [1]Martin Pruschy, [1]Gabriela Studer, [1]Jan Unkelbach, [1]Matthias Guckenberger, [3]Ender Konukoglu, [2]Philippe Lambin

[1] Department of Radiation Oncology, University Hospital Zurich, University of Zurich, Rämistrasse 100, 8091 Zürich, Switzerland

[2] Maastricht University Medical Centre - GROW-School for Oncology and Developmental Biology, Department of Radiation Oncology - MAASTRO, Dr. Tanslaan 12, 6229 Maastricht, The Netherlands

[3] Computer Vision Laboratory, ETH Zurich, Sternwartstrasse 7, 8092 Zurich, Switzerland

Published in the Radiotherapy and Oncology, 2017

**Abstract**

Purpose

This study investigated an association of post-radiochemotherapy (RCT) PET radiomics with local tumor control in head and neck squamous cell carcinoma (HNSCC) and evaluated the models against two radiomics software implementations.

Materials and methods

649 features, available in two radiomics implementations and based on the same definitions, were extracted from HNSCC primary tumor region in 18F-FDG PET scans 3 months post definitive RCT (training cohort n=128, validation cohort n=50) and compared using the intraclass correlation coefficient (ICC). Local recurrence models were trained, separately for both implementations, using principal component analysis (PCA) and the least absolute shrinkage and selection operator. The reproducibility of the concordance indexes (CI) in univariable Cox regression for features preselected in PCA and the final multivariable models was investigated using respective features from the other implementation.

Results

Only 80 PET radiomic features yielded ICC>0.8 in the comparison between the implementations. The change of implementation caused high variability of CI in the univariable analysis. Both final models performed equally well in the training and validation cohorts (CI>0.7) independent of radiomics implementation.

Conclusion

The two post-RCT PET radiomic models, based on two different software implementations, were prognostic for local tumor control in HNSCC. However, 88% of the features was not reproducible between the implementations.

**Keywords**: post-radiochemotherapy 18F-FDG PET, radiomics, local tumor control modelling, software implementation, reproducibility

## Introduction

Head and neck squamous cell carcinoma (HNSCC) is one of the most common cancers worldwide with tobacco and alcohol consumption as well as HPV infection being the important risk factors. The standard of care for patients with locally advanced HNSCC is definitive radiochemotherapy (RCT). The locoregional recurrence rate is high, exceeding 50% in HPV negative oropharyngeal carcinoma and non-oropharyngeal cancers [1, 2]. A meta-analysis of post-RCT 18F-fluorodeoxyglucose positron emission tomography (18F-FDG PET) studies reported sensitivity and specificity of around 80% in respect to detection of local tumor recurrence or persistence in HNSCC [3]. Additionally, post-RCT FDG PET has been shown to correlate with overall survival [4].

Radiomics, a high throughput method for quantification of medical images, has been shown a promising input for treatment response modelling [5-11]. It is based on a comprehensive and quantitative analysis of a region of interest performed on different levels: shape, intensity, texture and filter-based analysis. Radiomics is a rapidly growing field of research. However, the studies have been predominantly performed in independent single-institution settings and consequently, the importance of workflow standardization has been indicated [5, 6].

Radiomics analysis requires several image pre-processing steps such as region of interest segmentation and extraction as well as image interpolation and discretization. These steps together with image acquisition and reconstruction parameters may influence radiomic features and therefore interchangeability of derived models (i.e. radiomic signatures) [12, 13]. Many institutions use different software packages for the analysis, which are often in-house developed. Although the implementations are based on the same mathematical definitions, it is likely that they will produce different results due to differences in implementation of algorithms as well as pre-processing [13].

To base clinical decisions on a prognostic model, its validation is required [14, 15]. Several strategies, characterized by different strength, can be used. A cross-validation is often implemented as a first step, followed by temporal validation using data from the same institution but from a different period. Finally, to achieve an unbiased validation, an external validation in an independent dataset should be performed [16]. Most of the radiomics studies have used cross-validation to quantify model performance and so far only one model has been validated in an external and independent dataset [17, 18]. Validation is usually performed by the same research group, using the same tools and methodology. However, radiomic features have been shown to vary with image acquisition parameters, pre-processing and contouring [5, 6, 12] and (to our knowledge) none of the previously published studies investigated the reproducibility of a radiomics-based prognostic model in terms of radiomics software implementation.

This study hypothesize that the prognostic value of radiomic features is software implementation dependent. First, we investigated whether the 3 months post-RCT follow-up 18F-FDG PET radiomics is prognostic for tumor recurrence in HNSCC. Two independent models were trained using two independent radiomics implementations and their performance was validated in a separate dataset. Subsequently, the reproducibility of these models was evaluated when their respective radiomic features were calculated with an independent software implementation.

## Materials and methods

*Imaging protocol and studied population*

This retrospective analysis was approved by the local ethical commission. HNSCC patients treated with definitive radiochemotherapy were enrolled in the study (128 patients in the training and 50 patients in the validation cohort). The validation cohort consisted of patients treated in an institutional phase II prospective study (NCT01435252) with a standardized imaging protocol (the same slice thickness and reconstruction algorithm). Surgery or induction chemotherapy were exclusion criteria (biopsy allowed). The characteristic of the studied cohorts is presented in the Table 1. All patients underwent 18F-FDG PET/CT imaging prior to the treatment and 3 months after the end of the treatment as a standard follow-up examination. Depending on patient's body weight, an activity of 170-470 MBq of 18F-FDG was injected intravenously after the measurement of blood sugar level. The PET acquisition was preformed 60 minutes after 18F-FDG injection with a 3 minutes scanning time and 15 cm axial field-of-view at each bed position. Total acquisition time of the PET was 12-18 min. Images were reconstructed with an iterative algorithm (2D or 3D reconstruction in the training cohort and 3D reconstruction in the validation cohort) with an in-plane pixel size and the slice thickness of 2.73 – 5.47 mm and 3.27 – 4.25 mm, respectively. All data was acquired in the same center.

| | | **Training cohort** | **Validation cohort** |
|---|---|---|---|
| | **Total number of patients** | 128 | 50 |
| | **Median follow-up (months)** | 46 (3-156) | 16 (3-28) |
| | **Number of local recurrences** | 38 (30%) | 13 (26%) |
| **Tumor stage** | **T1/T2** | 43 (34%) | 6 (12%) |
| | **T3/T4** | 85 (66%) | 44 (88%) |
| **HPV status** | **Positive** | 31 (24%) | 22 (44%) |
| | **Negative** | 36 (28%) | 28 (66%) |
| | **Unknown** | 61 (48%) | 0 |
| **Tumor site** | **Oropharynx** | 91 (71%) | 29 (58%) |
| | **Hypopharynx** | 22 (17%) | 7 (14%) |
| | **Larynx** | 11 (9%) | 7 (14%) |
| | **Oral cavity** | 4 (3%) | 7 (14%) |
| **Treatment** | **Radiotherapy** | on average 70 Gy (68 – 72 Gy) | 70 Gy |
| | **Chemotherapy** | Cisplatin (40 mg/m2, up to 7 cycles) or cetuximab (loading dose 400 mg/m2 followed by 250 mg/m2 weekly) | Cisplatin/cetuximab (weekly same doses as in training cohort) with or without consolidation cetuximab (500 mg/m2 biweekly x 6) |
| **PET scanners** | **GE Discovery STE** | 64 (50%) | 39 (78%) |
| | **GE Discovery 690** | 10 (8%) | 6 (12%) |
| | **GE Discovery RX** | 23 (18%) | 5 (10%) |
| | **GE Discovery HR** | 15 (12%) | |
| | **GE Discovery LS** | 16 (12%) | |

*Table 1. Detailed characteristic of studied cohorts.*

*Image pre-processing and radiomics analysis*

Tumors were semi-automatically segmented in the pre-treatment PET scans using a gradient-based method implemented in MIMVISTA (MIM Software Inc., Cleveland, OH, USA). The pre-treatment and post-treatment scans were rigidly registered and contours were transferred to post-treatment scans. To account for differences in image reconstruction grid all scans were rescaled to 5.5 mm cubic voxels using linear interpolation. This corresponds to the smallest resolution in the studied dataset.

The pre-processed images were shared between the institutions. Post-RCT metabolic heterogeneity was studied in the region of the primary tumor (Figure 1). Two independent software implementations were used: implementation from the University Hospital Zurich (USZ) and the MAASTRO clinic (MAASTRO). In total 649 features, which were based on the same definition and available in both implementations, were extracted:

- Shape (n = 8)
- Intensity-based (n = 17)
- Texture: the Gray Level Co-occurrence Matrix (GLCM; n = 24), the Neighborhood Gray Tone Difference Matrix (NGLTDM; n = 4), the Gray-Level Size Zone Matrix (GLSZM; n = 14), the Gray-Level Run Length Matrix (GLRLM; n = 14).
- Filter-based: Wavelet coiflet (n = 568).

The full list of the extracted features is presented in the supplemental material. A bin size of 0.5 SUV was used for image intensity discretization. The consistency of radiomic features calculated in two different implementations was studied using the two-way mixed single measures intraclass correlation coefficient (ICC).

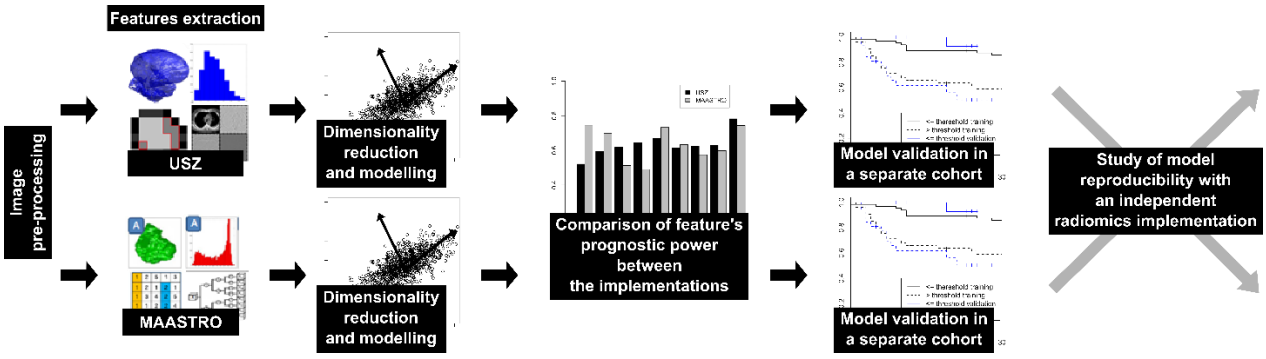


*Figure 1. Scheme of the reproducibility analysis of the local tumor control models using two independent radiomics implementation.*

*Features preselection and comparison of the features' prognostic power between the radiomics implementations*

The following feature selection procedure was used. First, a principal component analysis (PCA) was performed to account for inter-feature correlations. The number of retained components was adjusted to represent 95% of data variance. Next, for each principal component one feature was selected to represent

it. To that end we determined the feature that correlated the most (the largest Pearson correlation coefficient) with the principal component.

The prognostic power of radiomic features selected in different implementations was investigated in an univariable Cox regression. The models were fitted separately for the USZ and MAASTRO implementations. To quantify the discriminative power of different models the concordance indexes (CI) were calculated and compared between the implementations. The p-value from Cox regression was corrected for the multiple testing using false discovery rate (FDR) < 10% and the number of features defined in the PCA. The statistical analysis was performed in R (v. 3.2.3).

*Prediction of local tumor recurrence and model reproducibility between the implementations*

To train a final model for the association of the radiomic features derived from post-RCT PET with the likelihood of tumor recurrence, the least absolute shrinkage and selection operator (LASSO) (100 times 5-fold cross-validated) was used for variable selection in multivariable Cox regression. Only the features preselected in the PCA were used in the multivariable analysis. A random sampling with replacement was used to create a different training set in each of the LASSO iterations. In the final model we included only radiomic features with selection rate higher than 70% among all random training sets. Patients were stratified into low- and high-risk of recurrence groups based on a threshold from the receiver operating characteristic curve for local recurrence at 18 months. The threshold was selected to equate the level of sensitivity and specificity. The groups were compared using G-rho test (p-value < 0.05). Two models were trained separately, one on the USZ and one on the MAASTRO feature set. Both models were validated in the independent cohort of patients.

Each trained model, based on the features calculated in one implementation, was later evaluated by calculating its respective features with the other independent implementation (Figure 1). The regression coefficients of the Cox model and the stratification threshold were then fixed. Model performance was quantified using the concordance index (CI). Additionally, the calibration of the models was investigated by calculation of the calibration slope based on the prognostic index [19]. The calibration slope equals 1 evidences the same level of discrimination in the training and validation datasets. Finally, the correlation of hazards obtained with two implementations and the reproducibility of the patients risk group assignments were investigated.

**Results**

*Radiomic features reproducibility between the two implementations*

The intraclass correlation coefficient was used to investigate features reproducibility. Out of 649 features, 46 and 80 were characterized by an ICC greater than 0.9 and 0.8, respectively. These were mostly histogram-based (92% of the features in the studied group based on the ICC > 0.8) and texture-based (68%) features calculated on the non-transformed images. The shape features showed intermediate reproducibility (50%), whereas the biggest discrepancy was observed for the wavelet features (supplement Figure 1S). The wavelet features where high-pass filter was applied more than once were the least

reproducible. A translation of the 0.5 SUV bin size to the wavelet coefficients was different between the implementations (see supplement section Wavelet). It resulted in a different number of analyzed gray levels in the wavelet maps (supplement Figure 3S). Additionally, the MAASTRO implementation uses an undecimated transform, whereas the USZ implementation uses the decimated one. This influenced the resolution of the analyzed maps.

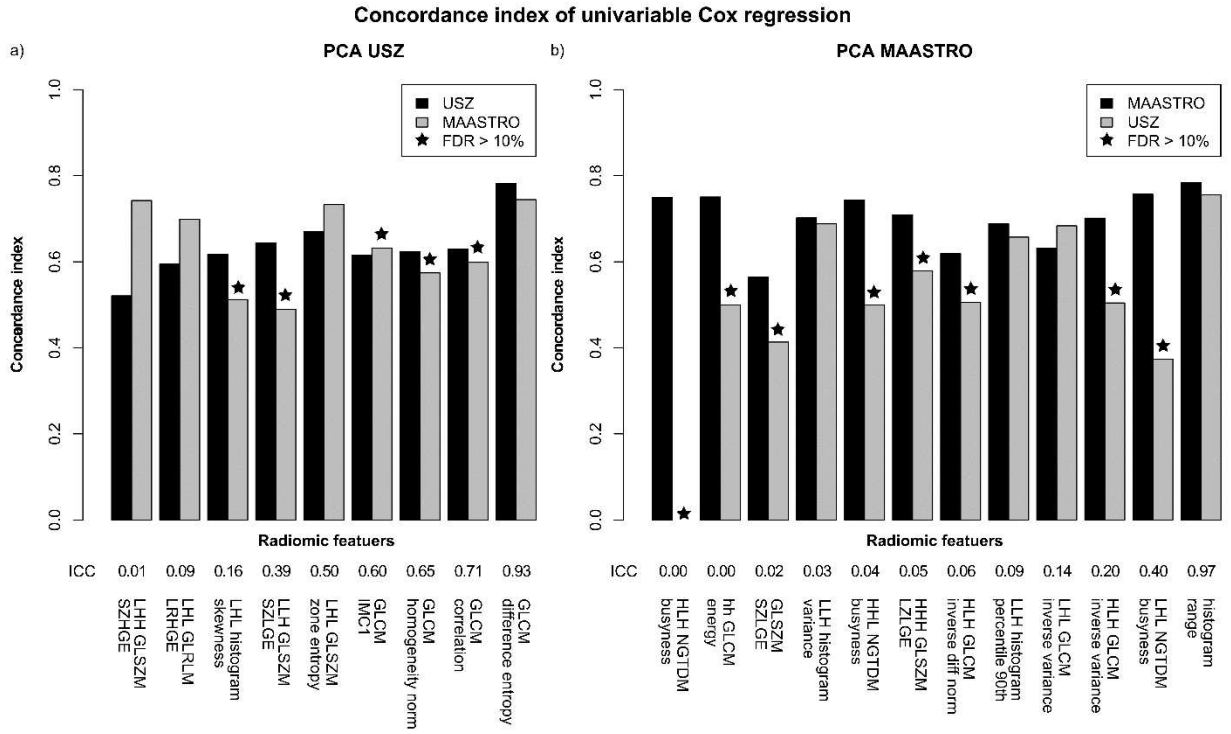


*Figure 2. Comparison of the features' prognostic power in the univariable Cox regression between the radiomics implementations. The fit was considered non-significant if false discovery rate (FDR) > 10%. The concordance indexes for the same feature varied between the radiomics implementations and this effect did not depend on the feature's intraclass correlation (ICC) from the implementations comparison. The LLL, LLH, LHL, LHH, HHH, HHL, HLH, HLL – denote the combination of wavelet filters in 3D (L – low-pass, H – high-pass).*

*Comparison of the features' prognostic power between the radiomics implementations*

In the principal component analysis, 31 and 33 components retained the 95% of data variance in the USZ and MAASTRO implementation, respectively. We found only 6 representative features based on the principal components analysis to be the same for both implementations. In a univariable Cox regression, 9 features in USZ and 12 features in MAASTRO implementation yielded a FDR < 10%. Among those features, more than 50% was not significant in the univariable Cox regression when calculated with secondary implementation (Figure 2). Even if the feature was significant in both implementations, a substantial difference in CI was observed. The features were grouped according to their FDR in the

secondary implementation (FDR < 10% or FDR ≥ 10%). No significant difference in the ICC values between those groups was observed (Wilcoxon test p-value > 0.05).

*Prediction of local tumor recurrence and comparison between the implementations*

In the multivariable analysis, GLCM difference entropy was found to be prognostic in the USZ implementation, whereas the histogram range was selected from the MAASTRO implementation. Radiomic features in the final local tumor recurrence models showed high level of reproducibility between the radiomics implementations (ICC > 0.9). A strong correlation (r > 0.9) between GLCM difference entropy and histogram range was observed independent of the implementation. There was a weak, significant correlation between selected radiomic features and tumor volume (r < 0.5).

Both models showed similar prognostic power in the training (5-fold cross-validation) and validation cohorts with CI ranging between 0.70 and 0.76 (Table 2) and allowed for a significant risk group stratification (Figure 3). In the validation cohort, the calibration slope was not significantly different from 1, indicating the preservation of model discriminative power (Table 2). Additionally, the models were prognostic in the group of HPV negative patients (supplement Figure 4S). In both models, tumors with higher risk of recurrence were characterized by a higher post-treatment metabolic heterogeneity (supplement Figure 5S).

| | | **Model developed using radiomic features from** | |
|---|---|---|---|
| | | **MAASTRO** | **USZ** |
| **Radiomic features** | | Histogram range | GLCM difference entropy |
| **Intraclass correlation** | | 0.97 | 0.93 |
| **Concordance Index** | | | |
| **MAASTRO features** | training | <u>0.76</u> | 0.75 |
| | validation | <u>0.73</u> | 0.73 |
| **USZ features** | training | 0.75 | <u>0.74</u> |
| | validation | 0.71 | <u>0.72</u> |
| **Calibration slope** | | | |
| **MAASTRO features** | | <u>1.20 (0.39 – 2.02)*</u> | 1.04 (0.27 – 1.95)* |
| **USZ features** | | 1.13 (0.39 – 1.88)* | <u>1.02 (0.20 - 1.83)*</u> |

*Table 2. Performance of PET radiomics models for prediction of local tumor control and the stability of radiomic features between two radiomics implementations (USZ and MAASTRO). Underlined values indicates results where the same implementation was used for the training of the model and model performance evaluation, * 95% confidence interval.*

The main research question asked in this work was to investigate the model performance when an independent radiomics implementation was used to calculate the hazards. Also in this case, the studied PET radiomics models achieved a very similar performance in terms of the concordance index as well as similar calibration slope (Table 2). It showed that the general discriminative power of the models was not affected by the change of the implementation. On the patient level, a strong correlation was observed between patient rankings based on the features from both implementations ($r > 0.9$). Most of the patients (around 90%) were correctly classified into low- or high-risk of recurrence group when the independent implementation was used (Figure 4).

**Discussion**

This study investigated the prognostic value of post-RCT PET radiomics in head and neck squamous cell carcinoma and tested the reproducibility of prognostic models between independent radiomics implementations (USZ or MAASTRO). Independent of the radiomics implementation used for model training, the prognostic model for local tumor control showed a good discriminative power with a concordance index higher than 0.7 in both training and validation cohorts. Both models significantly stratified patients into low- and high-risk of recurrence groups. Furthermore, the validation of the models using an independent radiomics implementation resulted in a similar concordance indexes. However, it is important to note that the reproducibility of the models is a consequence of the high ICC between the implementations for the selected features. In the modelling process we have observed that the discriminative power of single radiomic features preselected for the multivariable analysis depended on the radiomics implementation.

The value of post-treatment FDG-PET imaging for assessment of residual disease is currently unclear [3]. Recently, it has been shown in a prospective study that the positive findings on 3 months post-treatment FDG PET are a prognostic factor for overall survival and cancer-specific survival [4]. Additionally, our work shows that the heterogeneity of 3 months post-RCT FDG activity in the region of primary tumor is related to the risk of tumor recurrence. Higher histogram range (range of SUV in the region of primary tumor) and higher GLCM difference entropy corresponded to higher risk of tumor recurrence. We have further shown that these radiomics models can also significantly stratify the HPV negative patients, who belong to a group with a generally bad prognosis. Another study found that a pre-treatment FDG PET radiomics has a similar prognostic power to our post-treatment model [20].

Our prognostic models were trained on a heterogeneous dataset, different PET scanners and reconstruction algorithms were used. However, we were able to validate obtained results on a dataset with a standardized imaging protocol (the same slice thickness and reconstruction algorithm). Our findings should be further validated in datasets from other centers as the lack of calibration between different PET scanners can affect the performance of the models [21]. Additionally, we have defined our region on interest based on the pre-treatment PET images and propagated it to the post-treatment scan. The model reproducibility should be tested against different registration methods for propagation of the delineated tumor volume.

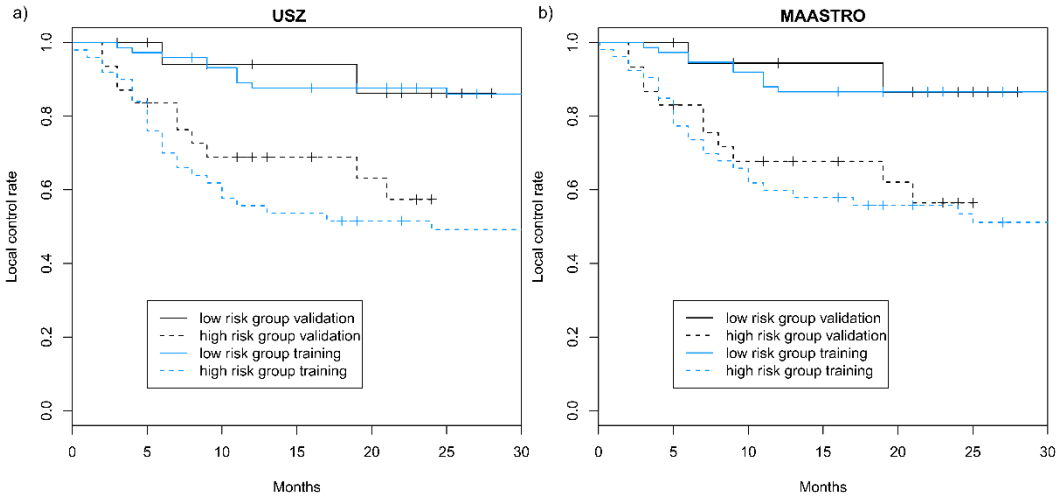


*Figure 3. PET radiomics-based local tumor recurrence models: a) USZ implementation, b) MAASTRO implementation. Local control rate curves split significantly (G-rho test p-value < 0.05) in both training and validation cohorts based on the optimal sensitivity-specificity thresholds at 18 months.*

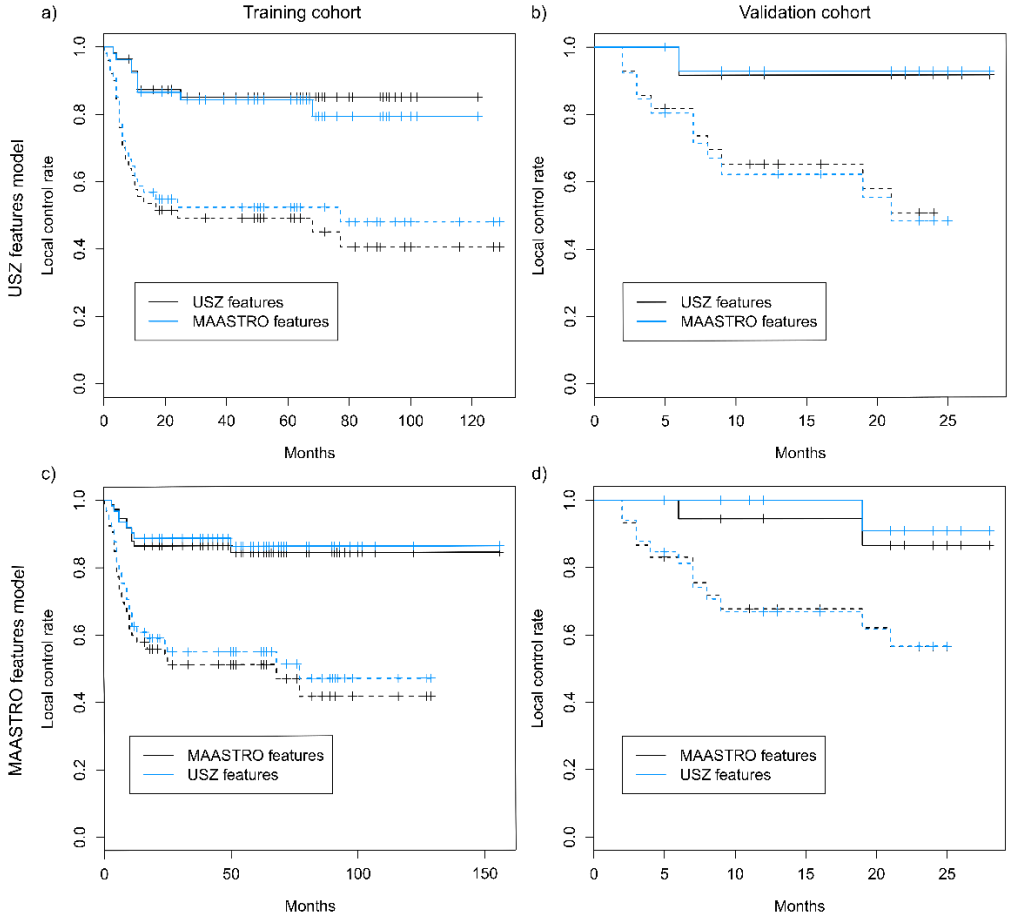


*Figure 4. Local control rate curves for low- and high-risk of recurrence groups based on the two PET radiomic models. The curves split significantly independent of the implementation (G-rho test p-value < 0.05).*

The two radiomics implementations used in this study are based on the same mathematical definition of radiomic features. Additionally, the image pre-processing (image and region of interest resizing) was performed independent of the radiomics implementation and the same bin size was used for image discretization [22]. Nevertheless, a relevant variability in radiomic features value was observed, mostly for the shape and wavelet features. It was most probably caused by differences in mask construction and wavelet transform workflow: especially in the translation of the bin size to the wavelet transformed images and the type of transform (decimated vs undecimated). For more details see supplement sections Contours mapping and Wavelet.. Variations of contour masks constructed from the same DICOM files is also a well-known issue in different treatment planning systems [23]. The comparison of the number of analyzed voxels, as well as minimum, maximum and mean value in the GTV between two radiomics implementations is shown in Figure 2S. The GTV constructed with USZ implementation was always larger then in MAASTRO implementation and consequently the minimum SUV in USZ implementation was always lower. Regarding wavelet transform, the two implementations transferred differently the bin size of 0.5 SUV into the wavelet coefficients space, which resulted in different number of analyzed gray levels (Figure S3). A separate study investigating a discriminative power of wavelet features obtained with the two gray level discretization methods could be conducted to clarify which method is more informative in the context of medical image analysis This study points out differences in radiomics workflow steps, which are rarely described in radiomics studies. Therefore, clear guidelines, such as the Image Biomarker Standardization Initiative [24], providing detailed description of radiomics workflow and implementation are needed. For a workflow comparison purpose, we are also open to share our source code upon request. Nonetheless, we showed that the majority of histogram and texture features was reproducible (i.e. high ICC values) despite the existing differences in contours mask construction. This result suggests that the mask creation is only a minor concern in the standardization aiming for the reproducibility of patients ranking and model prognostic power.

Our final prognostic models for local tumor recurrence were reproducible when the features from the independent radiomics implementation were used, which can be explained by the fact that both models consisted of radiomic features with a high ICC in the comparison between the implementations (ICC > 0.9). Most of the available radiomic features showed a much lower agreement in this comparison. A large variation in concordance indexes was observed for the radiomic features preselected in the principal component analysis. Most of the features preselected in one implementation were not significantly associated with local tumor recurrence in the other implementation in the univariable analysis. This shows, for the first time on a clinically relevant model and dataset, that a model developed by one institution should not be directly transferred to another center, which uses a different radiomics implementation, without rigorous comparison. We recommend that each model, additionally to a detailed description of the radiomics implementation, should be published with a sample dataset and corresponding radiomics signature, such as a recently published digital phantom [25]. This will allow for a comparison of results obtained from a model, before it will be used in a prospective cohort.

In conclusion, this study shows the potential of post-RCT FDG-PET radiomics for early identification of patients with a high risk of local tumor recurrence. It also raises an awareness of the impact of radiomics software implementation on model reproducibility.

**Conflict of Interest Statement:**

The authors declare that there is no conflict of interest.

**Acknowledgements:**

The project was supported by the Clinical Research Priority Program Tumor Oxygenation of the University of Zurich, by a grant of the Matching Fund of the University of Zurich. The clinical study used as validation dataset was supported by a research grant from Merck (Schweiz) AG. Additionally, authors acknowledge financial support from the ERC advanced grant (ERC-ADG-2015, n° 694812 - Hypoximmuno), the QuIC-ConCePT project (IMI JU; grant no. 115151). This research is also supported by the Dutch technology Foundation STW (grant n° 10696 DuCAT & n° P14-19 Radiomics STRaTegy), which is the applied science division of NWO, and the Technology Programme of the Ministry of Economic Affairs. Authors also acknowledge financial support from the EU 7th framework program (ARTFORCE - n° 257144, REQUITE - n° 601826), SME Phase 2 (EU proposal 673780 – RAIL), EUROSTARS (DART), the European Program H2020-2015-17 (BD2Decide - PHC30-689715 and ImmunoSABR - n° 733008), Interreg V-A Euregio Meuse-Rhine ("Euradiomics"), Alpe d'HuZes-KWF (DESIGN), Kankeronderzoekfonds Limburg from the Health Foundation Limburg, the Zuyderland-MAASTRO grant and the Dutch Cancer Society.

**References**

[1] Ang KK, Harris J, Wheeler R et al. Human papillomavirus and survival of patients with oropharyngeal cancer. N Engl J Med, 2010;363:24-35.

[2] Lassen P, Primdahl H, Johansen J. et al Impact of HPV-associated p16-expression on radiotherapy outcome in advanced oropharynx and non-oropharynx cancer. Radiother Oncol. 2014;113:310-316.

[3] Gupta T, Master Z, Kannan S et al. Diagnostic performance of post-treatment FDG PET or FDG PET/CT imaging in head and neck cancer: a systematic review and meta-analysis. Eur J Nucl Med Mol Imaging 2011;38:2083-2095.

[4] Kim SA, Roh JL, Kim JS et al. 18F-FDG PET/CT surveillance for the detection of recurrence in patients with head and neck cancer. Eur J Cancer 2016;72:62-70.

[5] Hatt M, Tixier F, Pierce L, Kinahan PE, Le Rest CC, Visvikis D. Characterization of PET/CT images using texture analysis: the past, the present… any future? Eur J Nucl Med Mol Imaging 2017;44;151-165.

[6] Yip SS, Aerts HJ. Applications and limitations of radiomics. Phys Med Biol 2016;61:150-166.

[7] Lambin P, Rios-Velazquez E, Leijenaar R et al. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer 2012;48:441-446.

[8] Bogowicz M, Riesterer O, Ikenberg K, et al. Computed Tomography Radiomics Predicts HPV Status and Local Tumor Control After Definitive Radiochemotherapy in Head and Neck Squamous Cell Carcinoma. Int J Rad Onc Biol Phys 2017; In press.

[9] Lambin P, Leijenaar RTH, Deist TM et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 2017; In press.

[10] Coroller TP, Grossmann P, Hou Y et al. CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. Radiother Oncol. 2015;114:345-50.

[11] van Timmeren JE, Leijenaar RTH, van Elmpt W et al. Survival prediction of non-small cell lung cancer patients using radiomics analyses of cone-beam CT images. Radiother Oncol. 2017;123:363-369

[12] Larue RT, Defraene G, De Ruysscher D, Lambin P, van Elmpt W. Quantitative radiomics studies for tissue characterization: A review of technology and methodological procedures. Br J Radiol 2017;90:20160665.

[13] Court LE, Fave X, Mackin D, Lee J, Yang J, Zhang L. Computational resources for radiomics. Translational Cancer Research 2016;5.

[14] Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Br J Cancer 2015;112:251-259.

[15] Rios Velazquez E, Hoebers F, Aerts HJ et al. Externally validated HPV-based prognostic nomogram for oropharyngeal carcinoma patients yields more accurate predictions than TNM staging. Radiother Oncol. 2014;113:324-330.

[16] Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research: validating a prognostic model. BMJ 2009;338.

[17] Leijenaar RT, Carvalho S, Hoebers FJ et al. External validation of a prognostic CT-based radiomic signature in oropharyngeal squamous cell carcinoma. Acta Oncol. 2015;54:1423-1429.

[18] Aerts HJWL, Rios Velazquez E, Leijenaar RTH et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun. 2014;5:4006.

[19] Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. BMC Medical Research Methodology, 2013;13.

[20] Bogowicz M, Riesterer O, Stark LS et al. Comparison of PET and CT radiomics for prediction of local tumor control in head and neck squamous cell carcinoma. Acta Oncol. 2017; In press

[21] Adams MC, Turkington TG, Wilson JM, Wong TZ. A systematic review of the factors affecting accuracy of SUV measurements. AJR Am J Roentgenol. 2010;195:310-320.

[22] Leijenaar R T H, Nalbantov G, Carvalho S et al. The effect of SUV discretization in quantitative FDG-PET Radiomics: the need for standardized methodology in tumor texture analysis. Sci Rep. 2015;5.

[23] Kulkarni BS, Sharma SD, Hansenet V et al. A prospective study of OAR volume variations between two different treatment planning systems in radiotherapy. International Journal of Cancer Therapy and Oncology 2015;3.

[24] Zwanenburg A, Leger S, Vallières M, Löck S, for the Image Biomarker Standardisation Initiative: Image Biomarker Standardisation Initiative. arXiv:161207003v4 2017.

[25] Lambin P. Radiomics Digital Phantom. Cancer Data 2016, DOI:10.17195/candat.2016.08.1.